An Empirical Application of Interactive Fixed Effect Model on Asset Pricing

Master's Thesis submitted

to

Prof. Dr. Wolfgang Karl Härdle

Humboldt-Universität zu Berlin

School of Business and Economics

Institute for Statistics and Econometrics

Ladislaus von Bortkiewicz Chair of Statistics

by

Bingling Wang

(586048)

in partial fulfillment of the requirements

for the degree of

Master of Science

Berlin, November 30, 2018

Acknowledgement

I would like to thank my thesis advisor Prof. Dr. Wolfgang K. Härdle at Institute for Statistics and Econometrics for his support and especially providing me the chance to present every progress of this paper in the institute meetings. I appreciate the amazing opportunity to discuss with professors and PhD students from different backgrounds and their inspiring comments and suggestions.

I would also like to thank Prof. Dr. Weining Wang for her advice on methodologies and her patient on my questions during the production of this thesis, as well as, the feedback and suggestions to conduct this research.

Finally, I must express my very gratitude to my parents for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis.

Abstract

Prediction of stock returns has been an important topic in both theoretical and empirical researches in Finance and statistics. The development of asset pricing models refers to the CAPM model, the arbitrage pricing theory, and the Fama-French multi-factor model. Connor et al. (2012) developed a nonlinear factor model which assumes that there is a nonlinear relationship between risk factors and stock returns, which can be a supplement of the linear multi-factor model, but the nonlinear effects are hard to explain with respect to economic meanings.

The interactive fixed effect model include not only a linear part which can be easily explained, but also a nonlinear part which explains the fixed effects in the residual of the linear estimation. This thesis aims to apply the interactive fixed effect model in asset pricing so as to improve both the linear and nonlinear asset pricing models. In addition, simulation and numerical application are included.

Key Words: Asset Pricing, Interactive fixed effect model, semi-parametric, factor models

Contents

List of Abbreviations

CAPM	Capital Asset Pricing Model
CRSP	Center for Research in Security Prices
PCA	principle componentsanalysis
PER	Percentiles
STD	Standard deviation

List of Figures

List of Tables

1 Introduction

Empirical studies show that individual stock returns have strong co-movement with financial and economic indicators, thus can be predicted cross-sectionally and over time.

Modern asset pricing theories committed to solve how the price of the asset is determined by risk, in the market equilibrium state. There are two leading asset pricing theories. The Capital Asset Pricing Theory developed by Sharpe (1964) and Lintner (1965) expresses a positive linear correlation between the expected return of an asset and the systematic risk of holding an asset, which is measured by the market β. Empirical tests on cross sectional CAPM such as Banz (1981), Rosenberg et al. (1985), and Basu (1983) shows the contradicted result to CAPM, that market βs are not sufficient to explain average returns. Firm Size, book to market equity and earnings-price ratios add the explanation of the average return. Fama and MacBeth (1973) found that this cross-sectional variation can be explained by two variables, firm size and book to market equity. Lettau and Ludvigson (2001) shows the relationship between average returns and market premium can be strong, weak or even disappear in different time periods. Extensions of CAPM with time varying covariates and Bollerslev et al. (1988) found that the conditional covariances of return of each asset are quite variable over time and are a significant determinant of time-varying risk premia. However, there is evidence that multivariate models should be considered.

Roll and Ross (1980) developed a more general asset pricing model that holds that the market has a factor structure. The expected return can be expressed as a linear function of the risk premium of K risk factors, which can be described as

$$y_i \quad = \mu_i + \textstyle\sum_{k=1}^{K} \beta_{ik} f_k + u_i \qquad i = 1, \dots N. \tag{1}$$

y_i is the return of the i-th asset, which has the expected value of μ_i. f_k is the k-th unobservable factor and β_{ik} is called factor loadings. u_i is the idiosyncratic risk of each asset with zero mean. Unobservable factors can be estimated by principle components analysis (PCA).

Empirical tests on this model include Fama and French (1993), which uses portfolio grouping to identify common risk factors in stock markets. They introduced a three-factor model: an overall market factor, firm size factor, and a firm value factor. Carhart (1997) estimated the excess returns of mutual funds by a four-factor model with additional factor of one-year momentum in stock returns. The more recent study of Fama and French (2016) developed a five-factor model that added a profitability and an investment factor to the three-factor model.

Connor and Linton (2007) introduced a new version of Fama-french three factor model by replacing the assumption of linear factor βs with an assumption that factor βs are smooth non-linear functions of observed security characteristics, and form factor-mimicking portfolios using nonparametric kernel methods. Connor et al. (2012) developed a weighted additive non-parametric factor model and estimated the factor returns and factor βs simultaneously without portfolio grouping. This model falls into the class of semi-parametric panel data models for large cross section and long time series, which can be denoted as:

$$y_{it} \quad = \textstyle\sum_{k=1}^{K} g_k\left(X_i\right) f_{kt} + u_{it} \qquad i = 1, \ldots N, k = 1, \ldots, K. \tag{2}$$

where y_{it} is the excess return to security i at time t, which can be explained by K factor returns at time t, denoted by f_{kt}. Factor loadings are denoted by $g_k\left(X_i\right)$. $g(\cdot)$ is an unknown function that map observable covariates to associated factor βs. $X_i = (X_{i1}, \ldots, X_{iq})$ are observable stock characteristics, and u_{it} are the mean-zero asset-specific returns.

Fan et al. (2016) improved Conner's semi-parametric panel data model by introducing loading coefficients expressed by $\lambda_{ik} = g_k\left(X_i\right) + \gamma_{ik}$, which relaxed the restriction in Conner's model that loading coefficients are fully explained by covariates. Here γ_{ik} is the component that cannot be explained by the covariates X_i. Thus the factor structure can be expressed as

$$y_{it} \quad = \textstyle\sum_{k=1}^{K} \left\{g_k\left(X_i\right) + \gamma_{ik}\right\} f_{kt} + u_{it} \qquad i = 1, \ldots N. \tag{3}$$

This method apply PCA on the projected data of Y on the sieve space spanned by $\{X_i\}_{i<=N}$. As long as the projection is genuine, the consistency of estimated factors and factor loadings requires only $N \longrightarrow \infty$ but T may or may not grow. This condition enables the estimation to be consistent in high-dimension-low-sample-size situation, which is attractive in the application in financial data.

In this paper, I combined this factor structure with a fixed effect panel data model by Bai (2009), which is a panel data model with unobservable multiple interactive effects, under both large N and T. The number of individuals N is assumed to grow without bound and can be larger than T. The interactive fixed effect model can be expressed as

$$y_{it} \quad = X_{it}'\beta + \lambda_i' f_t + \epsilon_{it} \qquad i = 1, \dots N; t = 1, \dots T. \tag{4}$$

X_{it} is a p dimensional vector which can be regarded as p sheets, each sheet is an $N \times T$ matrix. The estimation of βs can be considered as a least square estimator with general error terms (GLS). Factors f_t and factor loadings λ_i are estimated by projecting error term $W_i = Y_i - X_i\beta$ onto the space spanned by F. This estimator is shown to be consistent in a convergence rate of $\mathcal{O}_{\mathcal{P}}\left((NT)^{-\frac{1}{2}}\right)$.

1.1 Motivation

In this paper, I am interested in the relationship of the change of excess returns of the stocks and the changes of firm specific covariates of each stock in S&P 500 index. In order to study both the time and individual effects between variables, I observed the return and firm specific covariates of same individual stocks in S&P 500 index over time, thus obtain the panel data of excess returns and covariates. To make all the variables stationary, I apply the first order difference to the six covariates.

Figure 1 shows the cross-sectional relationship between excess returns on 2018-09-13 and the five factors of market capitalization, book to market ratios, earnings to price ratio, momentum and volatilities on the date 2018-09-06. The relationship between excess returns and the momentum factor seems to be linear, and that of the volatility factor seems to be linear with heteroscedasticities. Moreover, the relationship between the book to market value and excess returns seems to be non-linear with a convex shape. The other factors seems to be uncorrelated with excess returns. However, this relationship can be changing due to different market conditions. For example on the date 2009-07-10, which is shown in Figure 2, there is weaker correlation between excess returns and firm specific covariates such as book to market ratio and earnings to price ratio, but relatively stronger correlation with market factors such as momentum and volatilities, which may due to the impact of the financial crisis on the market.

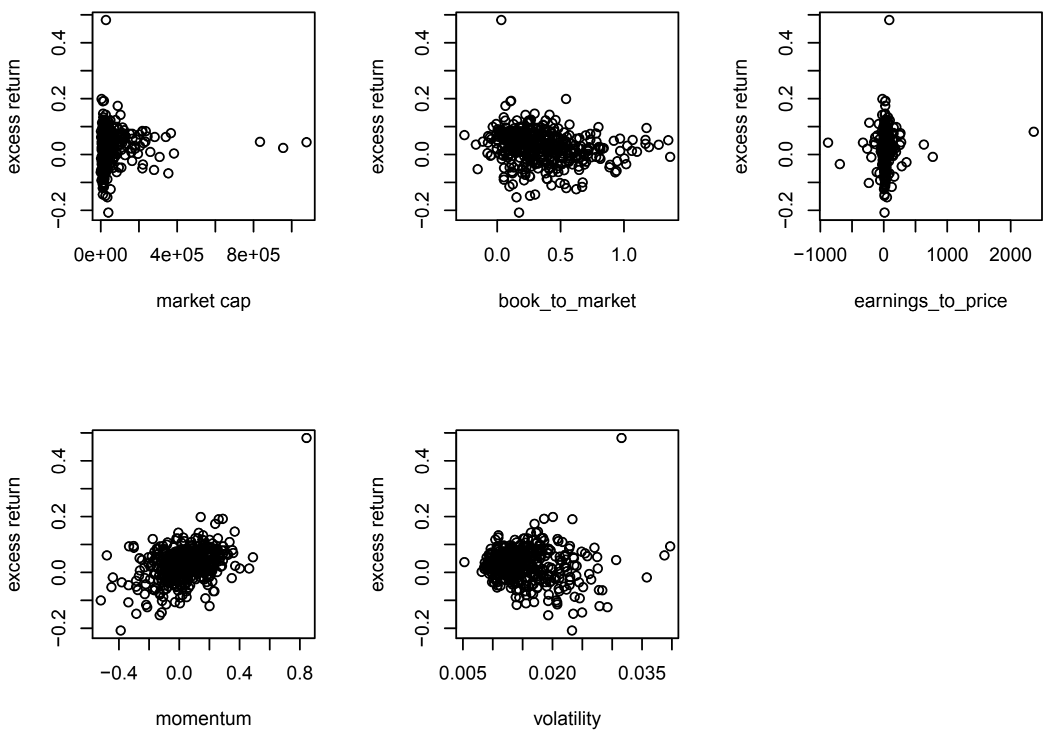

Figure 1: scatter plots of covariates and excess returns on 2018-09-06

IFEestimation

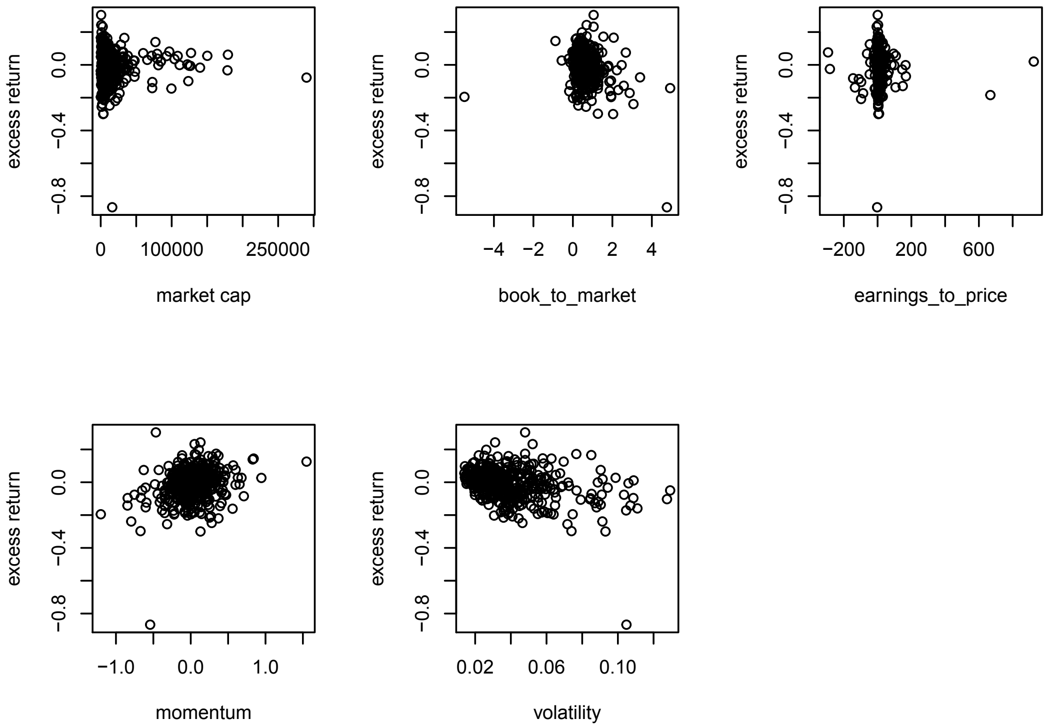

Figure 2: scatter plots of covariates and excess returns on 2009-07-10

IFEestimation

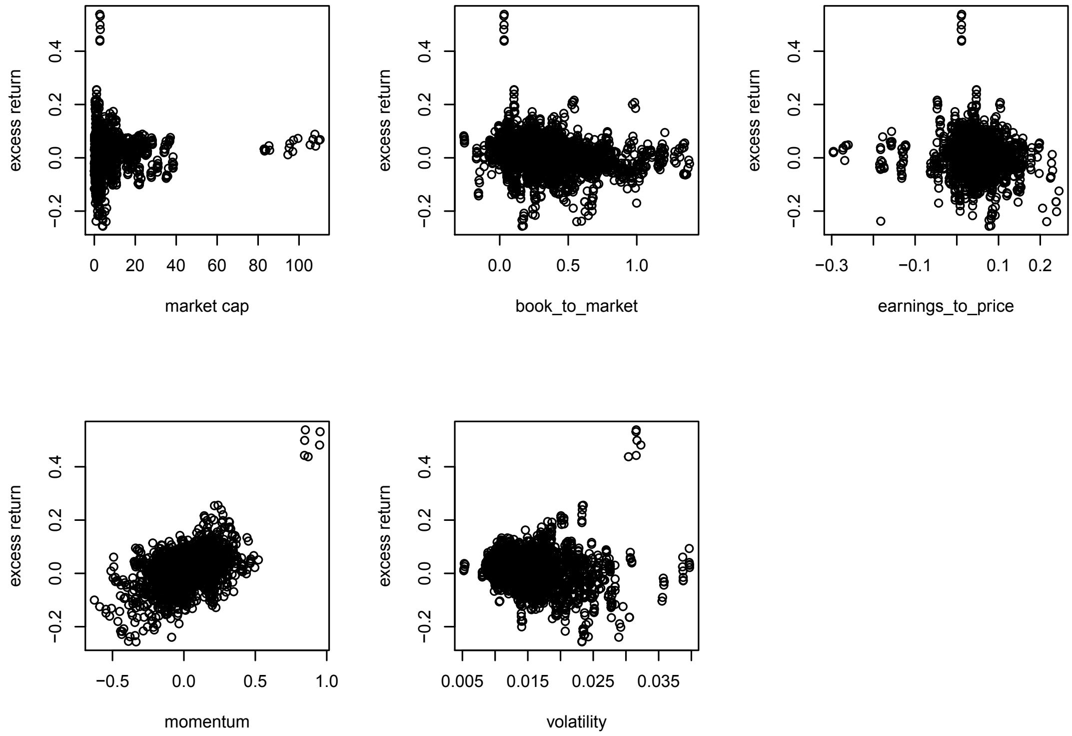

Figure 3: scatter plots of covariates and excess returns on 2018-09-28–2018-10-04

IFEestimation

Figure 3 and Figure 4 are the scatter plot of relationship between 5-trading-days panel excess returns and the previous 5-trading-days panel data of the five factors in 2009 and in 2018. With more samples over time, the relationship can be more consistent and linear relationship between excess returns and the changes of covariates still existed, but there are heterogeneity of intercepts which is related to both time effect and individual effect. Thus I consider panel data model with fixed effect.

I apply the same model as Bai (2009) in asset pricing case and using a projection based approach to capture both linear and non-linear correlations between the excess return of assets and the six risk factors mentioned in the papers of the Fama and French (1993), Carhart (1997) and Fama and French (2016). I tried to answer two research questions: Is Interactive Fixed Effect model works better to understand common dependence among multivariate stock market data better than conventional factor analysis? Will interactive fixed effect model improve the Fama-French linear

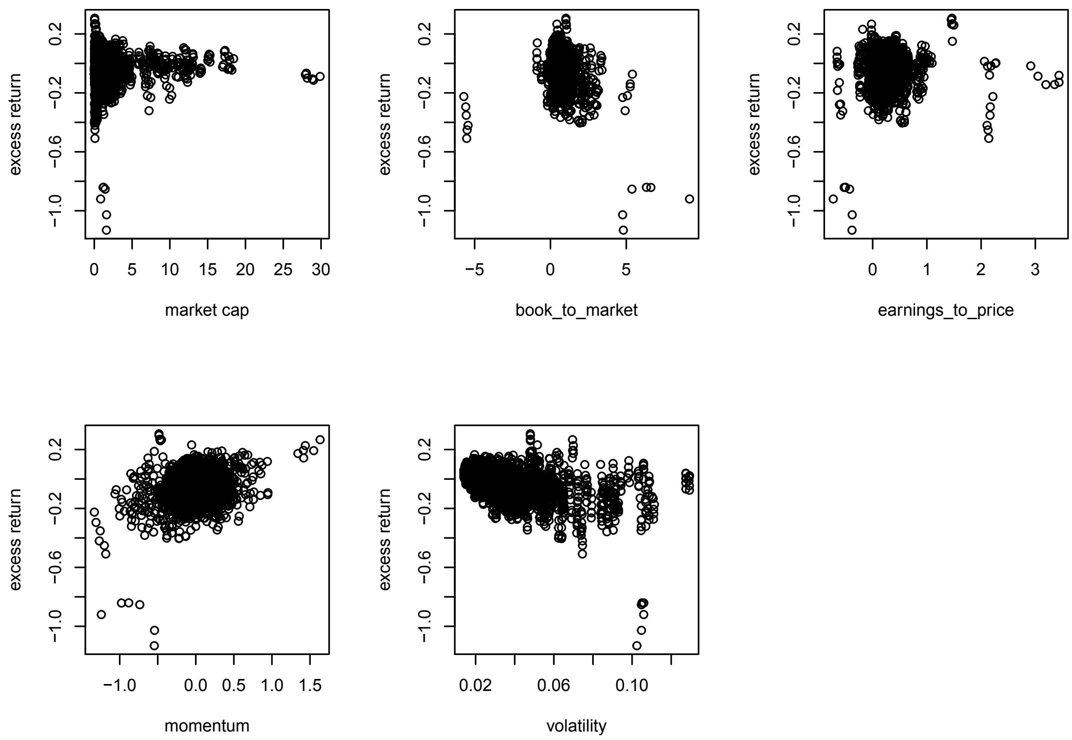

Figure 4: scatter plots of covariates and excess returns on 2009-07-01–2009-07-09
QIFEestimation

factor model?

The paper is organized as follows. The next section describes the interactive fixed effects model and the approximation methods I used to estimate it. Section 3 describes the data set and Section 4 presents the results. Finally, Section 5 concludes.

2 Model

2.1 Panel Model with Interactive Fixed Effects

Past empirical research on asset pricing model focused mainly on cross-sectional data or pooled cross-sectional data. In this paper we employ the panel asset pricing model in order to study more complicated behavior of stock returns, including time and individual heterogeneities.

Consider following Panel data model,

$$\begin{aligned} y_{it} &= x_{it}^\top \beta + \epsilon_{it}, \\ \epsilon_{it} &= \lambda_i^\top f_t + u_{it} \end{aligned} \qquad i = 1, \ldots, N \quad ; \quad t = 1, \ldots, T, \tag{5}$$

where y_{it} denotes the excess return of individual stock i in time t, which is the response variable we are interested in. x_{it} is a Q-dimensional vector of Q observable covariates related to stock i in time t. Covariates can be firm specific characteristics in Fama-French factor model, or macroeconomic indicators such as interest rate, exchange rate and long-term government bond interest rate, etc. β is a Q-dimensional vector of parameters which demonstrates the linear relationship between x_{it} and y_{it}.

The error term of the linear regression denoted by ϵ_{it} can be considered to have a factor structure, where number of factors is denoted by K. $\lambda_i^\top = (\lambda_{i1}, \ldots, \lambda_{iK})$ is the corresponding factor loadings associated with the k-th factor for individual i, and $f_t^\top = (f_{t1}, \ldots, f_{tK})$ is the unobserved factor returns of time t. And u_{it} represent the idiosyncratic return which cannot be explained by covariates and factors. u_{it} is assumed to be i.i.d and has finite variance σ_v^2. The factor structure can be reformulated as:

$$\epsilon_{it} = \sum_{k=1}^{K} \lambda_{ik} f_{kt} + u_{it} \qquad i = 1, \ldots, N \quad ; \quad t = 1, \ldots, T. \tag{6}$$

Given the parameter β, this model not only captures the linear relationship between risk factors and excess returns, but also allows the observable covariates x_{it} to be correlated with intercept f_t and λ_i, which can be suitable to deal with heteroscadasticity in an unknown form. Here λ_i describes the heterogeneity across individuals, which denotes all unobserved, time-constant effects that are not contained in x_{it}. f_t describes the unobserved common effects on individuals but change across time.

Conventional fixed effect model includes the individual effect and time effect in an additive form, which only captures the correlation between x_{it} and f_t or λ_i alone.

While interactive fixed effect model includes the cross-product term of f_t and λ_i, which captures in addition the collective effect of time and individual heterogeneity on x_{it}.

2.2 Estimation Procedure

Due to the interactive form of fixed effects, the standard panel data estimation method cannot be applied. In this paper, we follow the approach of Connor et al. (2012) and Fan et al. (2016), introduce a non-linear relationship between covariates and factor loadings. We start from projecting λ_i over the space spanned by $\overline{x}_{i\cdot}$, which can be expressed as

$$\lambda_i \quad = g\left(\overline{x}_{i\cdot}\right). \tag{7}$$

Here $\overline{x}_{i\cdot}$ denotes the average of x_{it} over time T. This non-linear relationship was proposed by Connor and Linton (2007) and developed to a more general form in the paper of Fan et al. (2016) with an error component γ_i. Thus we implement the following relationship

$$\lambda_i \quad = g\left(\overline{x}_{i\cdot}\right) + \gamma_i, \tag{8}$$

where $\gamma_i = \left(\gamma_{i1}, \ldots, \gamma_{iK}\right)^\top$ is a $K \times 1$ vector of the loading coefficients components that cannot be explained by the covariates $\overline{x}_{i\cdot}$. Assume γ_i has zero mean and is independent of $\overline{x}_{i\cdot}$ and u_{it}.

In other words, the whole model can be rewritten as

$$y_{it} \quad = x_{it}^\top \beta + \left\{g(\overline{x}_{i\cdot}) + \gamma_i\right\}^\top f_{kt} + u_{it}, \qquad i = 1, \ldots, N \quad ; \quad t = 1, \ldots, T. \tag{9}$$

If we assume f_t to be uncorrelated with x_{it}, the following composed error term is also uncorrelated with x_{it}

$$v_{it} \quad = \textstyle\sum_{k=1}^{K} \gamma_{ki} f_{kt} + u_{it}. \tag{10}$$

Thus consistent estimator β can be obtained with standard estimation approach for panel data model with random effects.

Rewrite the equation (9) into matrix form,

$$y_t \quad = X_t \beta + G(\overline{X}) f_t + v_t \qquad t = 1, \ldots, T, \tag{11}$$

where y_t and v_t are the $K \times 1$ vectors of y_{it} and v_{it}, X_t is the $N \times Q$ matrix of x_{it}, β is the $Q \times 1$ unknown parameter to estimate. f_t is the $K \times 1$ vector of f_{tk}, and $G(\overline{X})$ is the $N \times K$ matrix of $g_k(\overline{x}_{i\cdot})$.

To estimate the parameter β, we need first estimate factor f_t and loading function $G(\overline{X})$. A common procedure is to project the error term of linear regression denoted by $(y_t - X_t\beta)$ on the space spanned by f_t. In this paper we consider a sieve estimation for $G(\overline{X})$ to estimate β and project on the space spanned by sieve basis functions of $\overline{X}$.

We suppose loading function $g(\cdot)$ can be approximated by some spline function. Given $\overline{x}_{i\cdot}$ is Q dimensional, assume for each k, $g_k(\cdot)$ is an additive function of Q variates,

$$g_k(\overline{x}_{i\cdot}) = \sum_{q=1}^{Q} g_{kq}(\overline{x}_{i\cdot,q}), \qquad k = 1, \dots K \quad , \quad q = 1, \dots Q \tag{12}$$

For each k, q, $g_{kq}(\cdot)$ can be obtained by sieve approximation expressed as below,

$$g_{kq}(\overline{x}_{i\cdot\cdot,q}) = \sum_{l=1}^{J} b_{l,kq}\phi_l(\overline{x}_{i\cdot\cdot,q}) + R_{kq}(\overline{x}_{i\cdot\cdot,q}) \qquad k = 1, \dots K \quad , \quad q = 1, \dots Q \tag{13}$$

where $\phi_\ell(\cdot)$'s are the spline basis functions. For $b_{\ell,kq}$'s are the sieve coefficients of the lth additive component of $g_k(\overline{x}_{i\cdot,q})$ corresponding to the kth factor loading, and R_{kq} is a "remaining function" that represents the approximation error. Also, J denotes the number of sieve terms which grows slowly as $N \to \infty$. We take the basic assumption for sieve approximation that $sup_x|R_{kq}(\overline{x}_i, q)| \to 0$, as $J \to \infty$.

Define

$$b_k^\top = (b_{1,k1}, \dots, b_{J,k1}, \dots, b_{J,kQ}, \dots, b_{J,KQ}) \in \mathbb{R}^{JQ},$$

$$\phi(\overline{x}_{i\cdot})^\top = (\phi_1(\overline{x}_{i\cdot,1}), \dots, \phi_J(\overline{x}_{i\cdot,1}), \dots, \phi_1(\overline{x}_{i\cdot,J}), \dots, \phi_Q(\overline{x}_{i\cdot,J})) \in \mathbb{R}^{JQ},$$

we can simplify equation (13) as

$$g_k(\overline{x}_{i\cdot}) = \phi(\overline{x}_{i\cdot})^\top b_k + R_k(\overline{x}_{i\cdot}). \tag{14}$$

Rewriting (14) in matrix form we obtain

$$G(\overline{X}) = \Phi(\overline{X})B + R(\overline{X}), \tag{15}$$

where $\Phi(\overline{X}) = (\phi(\overline{x}_{1\cdot}), \ldots, \phi(\overline{x}_{N\cdot}))^\top$ is a $N \times JQ$ matrix of basis functions, $B = (b_1, \ldots, b_k)$ is a $JQ \times K$ matrix of sieve coefficients, and $R(\overline{X})$ is a $N \times K$ matrix with the (i,k)th element of $\sum_{q=1}^{Q} R_{kq}(\overline{x}_{i\cdot,q})$.

Then, substituting (15) into (11) we obtain

$$y_t = X_t\beta + \Phi(\overline{X})Bf_t + v_t + R(\overline{X})f_t, \qquad t = 1, \ldots, T. \tag{16}$$

The residual term of this regression model consists of two parts: the sieve approximation error $R(\overline{X})f_t$ and the idiosyncratic error v_t that is of the form $v_t = \Gamma f_t + u_t$, where Γ is a $N \times K$ matrix of unknown loading coefficients.

With the aim of estimating β, we define P_Φ as the projection matrix onto $\mathcal{X}$, where $\mathcal{X}$ is the sieve spaced spanned by the basis functions of $\overline{X}$. More precisely, P_Φ is the $N \times N$ projection matrix of the form

$$P_\Phi = \Phi(\overline{X})(\Phi(\overline{X})^\top\Phi(\overline{X}))^{-1}\Phi(\overline{X})^\top. \tag{17}$$

Then, multiplying $(I_N - P_\Phi)$ on both sides of equation (16), we have

$$(I_N - P_\Phi)y_t = (I_N - P_\Phi)X_t\beta + (I_N - P_\Phi)v_t, \qquad t = 1, \ldots, T. \tag{18}$$

Therefore, one can obtain the estimation of β by partialling out the effect of factors f_t,

$$\hat{\beta} = \left(\sum_{t=1}^{T} X_t^\top(I_N - P_\Phi)X_t\right)^{-1} \sum_{t=1}^{T} X_t^\top(I_N - P_\Phi)y_t, \tag{19}$$

where $X_t^\top(I_N - P_\Phi)X_t$ is assumed to be asymptotically nonsingular.

2.3 B-Spline Estimation

In this section, I will give a more detailed introduction to B-Spline Regression. B-spline, also called Basis splines, is a special representation of spline. It is a linear combination of B-spline curve, which can be used to fit the data.

Given $m+1$ nodes t_i, which is in the interval $[0,1]$, and satisfy $t_0 < t_1 < \ldots < t_m$. Thus an n-degree B-spline is a parametric curve which consists of basis Spline to the

$$\mathbf{S} : [0,1] \to \mathbb{R}^2$$

$$\mathbf{S}(t)=\sum_{i=0}^{m}\mathbf{P}_i b_{i,n}(t)\ ,\ t\in[0,1]$$

P_i is called a control point or a de Boor point. m+1 n-degree B-spline base can be defined by the Cox-de Boor recursive formula

$$b_{j,0}(t):=\begin{cases}1 & t_j<t<t_{j+1}\\ 0 & \dots\end{cases}$$

$$b_{j,n}(t):=\frac{t-t_j}{t_{j+n}-t_j}b_{j,n-1}(t)+\frac{t_{j+n+1}-t}{t_{j+n+1}-t_{j+1}}b_{j+1,n-1}(t).$$

2.4 Significance Test

We consider the following significance test on the linear coefficients,

$$H_0:\hat{\beta}=0 \qquad H_1:\hat{\beta}\neq 0$$

This motivates a t-test according to asymptotic distribution of $\hat{\beta}$. We normalize the test statistic by its asymptotic variance leads to the test statistic

$$S_t=\sqrt{NT}Diag\left(\widetilde{V}\right)^{-\frac{1}{2}}\left(\hat{\beta}-\beta\right)\xrightarrow{d}N\left(0,1\right)$$

where

$$\widetilde{V}=\widetilde{V}_{\pi}^{-1}\widetilde{V}_{\Gamma,u}\widetilde{V}_{\pi}^{-1} \tag{20}$$

$$\widetilde{V}_{\pi}=\lim_{N,T\to\infty}\frac{1}{NT}\sum_i\sum_t\mathrm{E}\left(\pi_{it}\pi_{it}^{\top}\right)$$

$$\widetilde{V}_{\Gamma,u}=\lim_{N,T\to\infty}\frac{\sigma_{\pi}^2}{N}\left(tr\left\{\mathrm{E}\left(\Gamma\Gamma^{\top}\right)\right\}+\frac{1}{T}\sum_t tr\left\{\mathrm{E}\left(uu^{\top}\right)\right\}\right)$$

And $\pi_{itq}s'$ are random variables with zero mean.

3 Data

I apply the method to real data of US stock market by collecting the stock daily data in S$P 500 index from 2018-07-01 to 2018-10-08. To ensure a sample size that is large enough, I use daily observed monthly excess return, which is collected daily of the cumulative excess return of the previous 25 days. The excess return is calculated by subtracting risk free return from the return of the stocks. Where daily risk free rate can be approximated with 13-week treasury bill daily rate. The data is obtained from CRSP. Daily stock price and firm specific characteristics are collected from compustat. Data are cleaned by removing the stocks with extreme values and null values, and end up with 447 stocks. For each stock, we consider 5 characteristics as in Fama and French (2016) and Carhart (1997), which are size, value, profitability, momentum and volatility. Size effect is measured by the market capitalization. Value and profitability characteristic is measured by book to market ratio, and price to earnings ratio. Momentum characteristic is calculated by the log return of cumulative return of the previous 126 trading days before the observation date, which is considered as the cumulative half-year return. And volatility characteristic is the standard deviation of the daily returns of the previous 126 trading days. the data description is demonstrated in Table 1.

Table 1: Some descriptive statistics of location and dispersion for 2582 observed data for the period from 2018-07-01 to 2018-10-08

variables	mean	median	STD	25th PER	75th PER	skewness	kurtosis
excess return	0.0085	0.0149	0.1041	-0.0350	0.0600	-1.3038	20.6208
size	2.9503	1.3131	5.2375	0.7251	2.8019	5.7829	58.0119
value	0.4361	0.3520	0.4491	0.2026	0.5721	22.3222	1844.1724
profitability	0.0951	0.0685	0.1467	0.0339	0.1230	6.9613	136.8922
momentum	0.0411	0.0625	0.2387	-0.0523	0.1644	-1.5853	14.9388
volatility	0.0192	0.0156	0.0123	0.0121	0.0218	3.2562	19.4698

4 Results

4.1 Simulation Results

In this section, we evaluate our estimation's performance via a simulation study. The simulation study can be divided into three steps:

i) Simulate the variables and coefficients.

ii) Estimate the parameter β and factor f_t.

iii) Compare the estimated results with the simulation setups.

Step i) : Starting from the following panel data model, see equation (5). Firstly, we simulate the factor model as equation (6). To reduce the dimension of factor matrices, we set the factor number K=3. Assume f_t is a strictly stationary process and $f_t \sim VAR(0, \Sigma_f)$ and fits the VAR(1) model:

$$f_t = \mu + \Pi f_{t-1} + e_t \qquad t = 1, \ldots, T, \tag{21}$$

Let f_t be the $K \times 1$ vector of f_{tk}, e_t be $K \times 1$ vector of error terms, μ be $K \times 1$ vector of 0, then Π is a $K \times K$ dimensional matrix of coefficients. In addition we set the covariance matrix $\Sigma_e = [0.9076, 0.0049, 0.0230; 0.0049, 0.8737, 0.0403; 0.0230, 0.0403, 0.9266]$, where $e_t \sim N(0, \Sigma_f)$. Then we generate a random $\Pi \in (-1, 1)$, i.e. $\Pi = [-0.0371, -0.1226, -0.1130; -0.2339, 0.1060, -0.2793; 0.2803, -.0755, -0.0529]$. For $t = 1$ we set the initial value f_1 by creating one sample from the multivariate normal distribution $\mathcal{N}(0, \mathbb{V}[f_t])$, where $\mathbb{V}[f_t] = [0.9371, 0.0330, 0.0266; 0.0330, 1.0176, -0.0148; 0.0266, -0.0148, 1.0065]$, which is calculated by $\mathbb{V}[F_t] = (I_{K \times K} - \Pi \otimes \Pi)^{-1}\text{vec}(\Sigma_f)$. For $t = 2, 3, \ldots, T$, f_t can be calculated by iteration.

The covariates X_t can be simulated by generate Q random samples of different multivariate distribution. Since the actual data is leptokurtic distributed, we generate two sets of values, one for a normal distribution and the other a t-distribution. We set $Q = 6$, $T = 3, 5, 10, 50, 126, 400$, and $N = 3, 5, 10, 50, 126, 400$.

To estimate the factor loadings, we follow equation (8), where $g(\cdot)$ is a $K \times 1$ vector of unknown functions which can be approximated by sieve regression with B-splines, as in equation (13). To simulate $\phi(\bar{x}_{i\cdot})$, we generate the B-spline basis matrix on each

$\bar{x}_{i\cdot}$. Let the degree of B-spline equal to n. The inner knots equal to the quantiles of $\bar{x}_{i\cdot}$, boundary knots equal to the largest and smallest value of $\bar{x}_{i\cdot}$. Note that in R the entire set of knots are obtained by adding $(n+1)$ lower boundary knot and $(n+1)$ upper boundary knot with the inner knots. So the total number of knots denoted by m is $(3n+2)$. For the each covariate, the number of basis function is $(m-n-1)$. We store the result of $\phi(\overline{x}_{i\cdot})$ in an $(N \times JQ)$ matrix denoted by $\Phi(\overline{X})$. Then we simulate the sieve coefficients. b_k contains J sieve coefficients corresponding to the k-th factor loading and the q-th variate. We simulate it by generating a $(JQ \times K)$ matrix range from 0.00001 to 0.01. For each factor, there are J coefficients corresponding to J basis functions. Here we denote this matrix by B. The approximation error $R_k(\overline{x}_{i\cdot})$ is then simulated by generate an $(N \times K)$ matrix of random samples from multivariate normal distribution $\mathcal{N}(0, 0.05)$. Therefore, according to the matrix form of equation (15), We can calculate $G(\overline{X})$.

At last, by the formula of equation (16), we obtain the simulated y_t. The N dimensional vector of composed error term v_t can be calculated as in equation 10, where $u_{it} \sim \mathrm{N}(0, 0.5)$, and $\gamma_{it} \sim \mathrm{N}(0, 0.0027)$ are simulated accordingly.

Step ii) : Set a $Q \times 1$ vector of β as [0.056,0.785,0.103,-0.087,0.914,-0.093], where $Q = 6$. According to the estimating method in part 2, the estimated coefficients $\hat{\beta}$ can be obtained by equation 19, where P_Φ is in the form of equation 17.

Then we estimate the K factors by obtain the first K eigenvectors of $\hat{E}^\top P_\Phi \hat{E}$. Where $\hat{E} = y_t - X_t\hat{\beta}$. The estimated coefficient matrix $\hat{B} = \frac{1}{T}\left[\Phi(\overline{X})^\top \Phi(\overline{X})\right]^{-1} \Phi(\overline{X})^\top \hat{E}\hat{f}_t$.

With all the estimation results above, we can obtain the prediction result that $\hat{y}_t = X_t\hat{\beta} + \hat{G}(\overline{X})\hat{f}_t$.

Step iii) : In Table 2, We report the root-mean-square error (RMSE) of the estimator $\hat{\beta}$ and factor f_t as well as predicted error of y_t related to two sets of X_t. It can be shown that our method has moderate levels of estimation accuracy in both scenarios considered.

For each group of X_t, no matter the distribution is leptokurtic or not, the estimation of β is becoming more accurate when N and T increase simultaneously, and the RMSE of β decreases faster when T is increasing than N is increasing, which means the RMSE of β is more sensitive to large T than large N. However, the situation is more complicated regarding to the estimation of f_t and y_t. For both sets of X_t, the estimation accuracy of factors seems to be uncorrelated with the dimension N or sample size T. The prediction error of y_t is becoming larger with the increase of dimension, but the increasing of sample size cannot necessarily increase the prediction accuracy.

In the case of large N, For linear coefficient β, the t-distributed X_t seems to have better estimation accuracy than normally distributed X_t. But the normally distributed data has smaller error in estimation of factors and the predictions.

Table 2: Monte Carlo Simulation results of 1000 iterations

IFEsimulation

X_dist		Normal distribution			t-distribution		
T	N	RMSE_β	RMSE_f	RMSE_y	RMSE_β	RMSE_f	RMSE_y
3	400	0.0768	0.8703	0.7419	0.0905	1.2278	0.7584
5	400	0.0856	0.9205	0.7158	0.0730	0.7514	0.6979
10	400	0.0422	1.3262	0.7699	0.0606	1.1143	0.7648
50	400	0.0202	0.7821	0.9784	0.0198	1.3302	0.9569
126	400	0.0144	1.5202	0.6914	0.0158	1.3306	0.6897
400	400	0.0099	1.2846	0.7045	0.0068	1.1489	0.7130
400	3	0.0779	1.4172	0.0000	0.2860	1.4489	0.0000
400	5	0.1046	1.4387	0.2363	0.0681	1.4410	0.2451
400	10	0.0430	1.6313	0.6084	0.0312	1.6314	0.5948
400	50	0.0562	1.4604	0.5608	0.0493	1.0353	0.5626
400	126	0.0165	1.6309	0.6897	0.0117	1.5712	0.7040

In Table 3, We report the same estimated and predicted error related to different sieve dimensions. We can conclude form the table that the larger the sieve dimension, the faster the estimated factor converging to the true value when N or T increase simultaneously or individually. The same results is also showed in the prediction. In addition, although the larger the sieve dimension, the better the spline fitted the curve, we should be aware of over-fitting.

Table 3: Monte Carlo Simulation results of 1000 iterations

QIFEsimulation

sieve dimension			J = 8			J = 32	
T	N	RMSE_beta	RMSE_f	RMSE_Y	RMSE_beta	RMSE_f	RMSE_Y
3	400	0.0754	0.8380	0.7477	0.1304	0.7318	0.6944
5	400	0.0818	1.2359	0.7377	0.1045	1.3544	0.7148
10	400	0.0421	1.0606	0.7687	0.0454	1.2065	0.7513
50	400	0.0202	1.4969	0.9824	0.0259	1.7401	0.9427
126	400	0.0145	0.6418	0.6902	0.0149	0.7114	0.6899
400	400	0.0092	1.4402	0.7060	0.0111	1.3747	0.6952
400	3	0.0753	1.4180	0.0000	0.0892	1.4154	0.0000
400	5	0.0995	1.4392	0.2352	0.0843	1.4394	0.2438
400	10	0.0423	1.6322	0.6071	0.0406	1.5535	0.6091
400	50	0.0466	1.8308	0.5527	0.0840	1.4491	0.5649
400	126	0.0159	1.6433	0.6896	0.0141	1.4450	0.6701

4.2 Application Results

In this section, moving window estimation is completed on the real data and the results are plotted below. We estimate the linear and non-linear impact of the five firm specific factors on the stock excess returns in the next window period, and conduct a significance test on the linear coefficient.

As one can see from the figures (5) to (9), the estimated linear coefficient of the five factors with a moving window of the size 10, 126, 400 trading days are plotted below. For each covariate, we assume a three factor structure and the factor loadings are estimated by sieve estimation with the dimension of 11. We can see the smaller the moving window size, the larger the linear coefficient $\hat{\beta}$, and the more frequently changed $\hat{\beta}$ is. This might be due to the shorter the observation window, the more sensitive that $\hat{\beta}$ is.

However, from figure (10) to (14), we can observe that the larger the moving window size, the more likely the linear coefficient to be significant. Thus, when choosing the window size, is a trade-off between significant linear coefficient and the actual linear impact of five characters on the return.

In figure (15) and (16) the estimated additive non-linear function for each factor is plotted. It is shown that the nonlinear part for different moving windows are different in shapes and are obviously nonlinear, which adds validation to our model to apply on the stock data.

Figure (17) helps us to evaluate the prediction and better understand the intuition of the model. we can see the prediction results are pretty good relative to the real excess returns. The linear part captures the changing in mean of excess returns, while the nonlinear part captures the fluctuations around the mean.

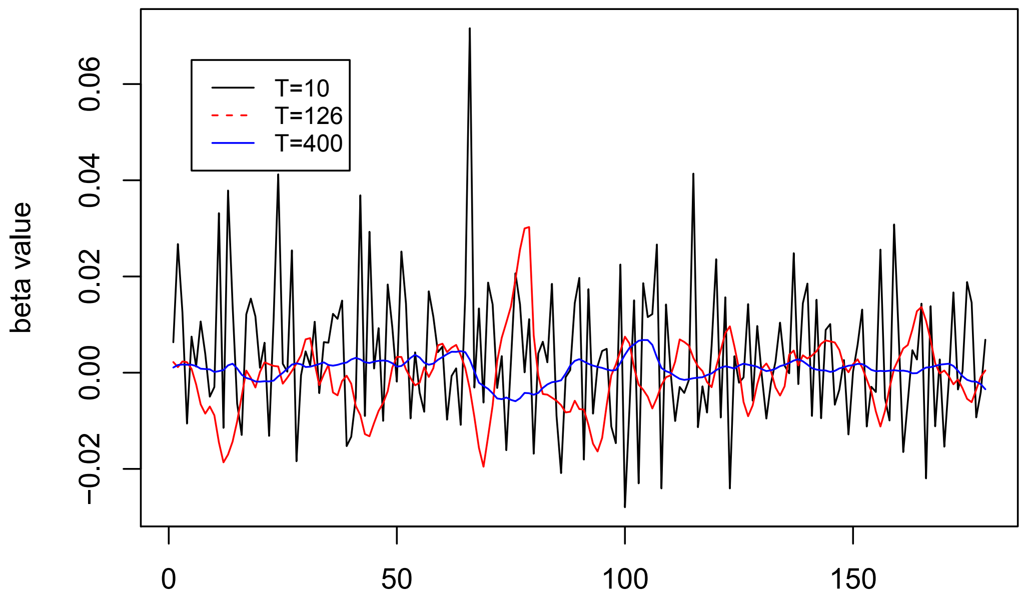

Figure 5: $\hat{\beta}$ results of size factor

IFEmoving_window

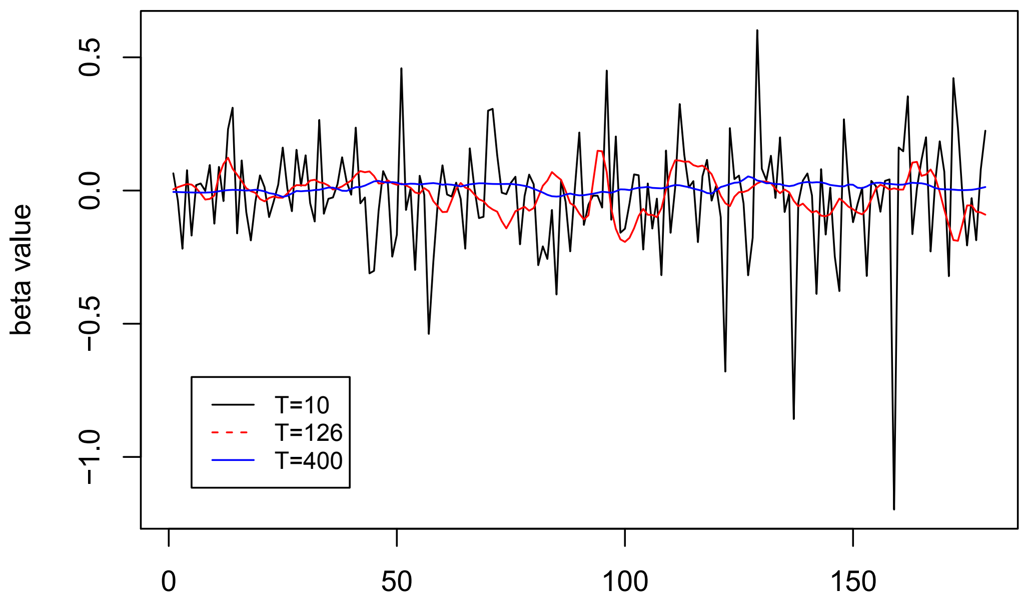

Figure 6: $\hat{\beta}$ results of value factor

IFEmoving_window

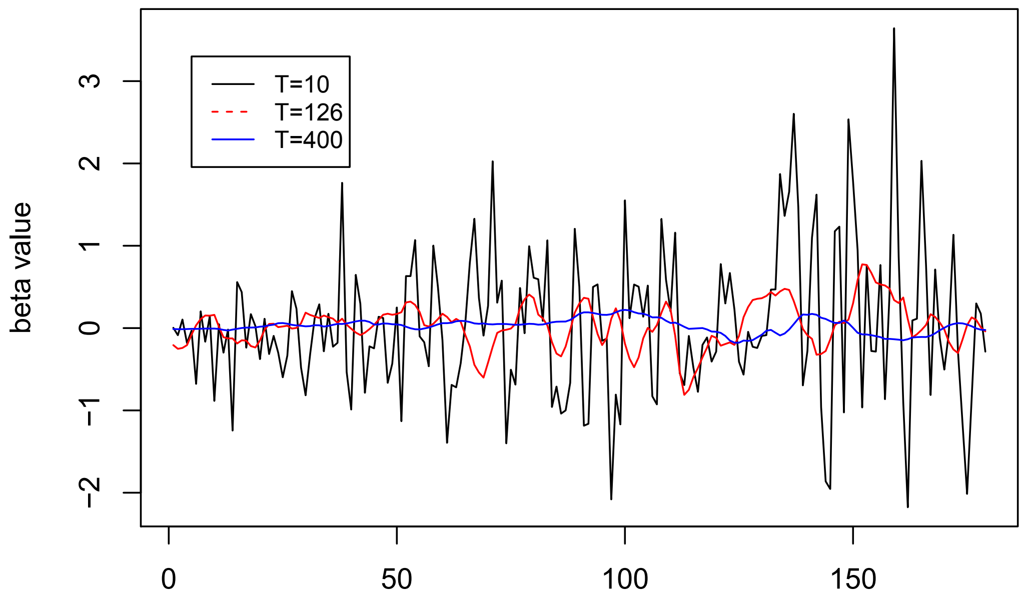

Figure 7: $\hat{\beta}$ results of profitability factor

IFEmoving_window

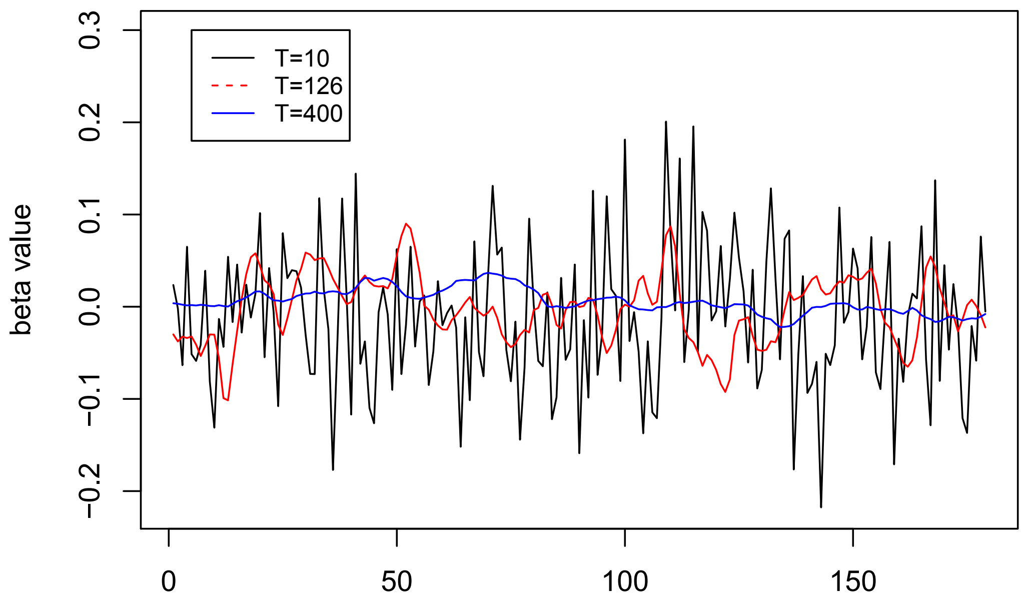

Figure 8: $\hat{\beta}$ results of momentum factor

IFEmoving_window

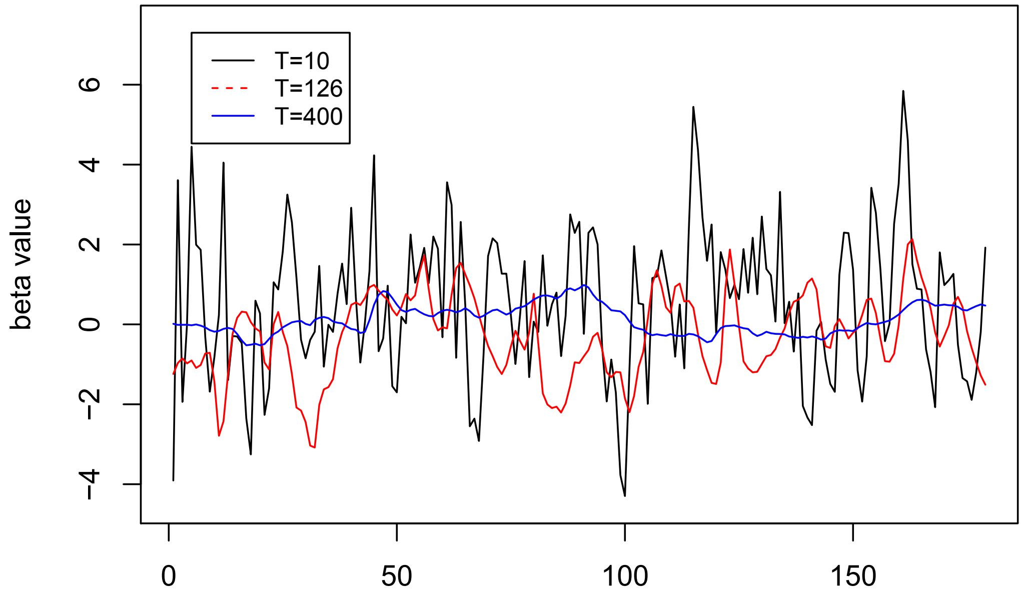

Figure 9: $\hat{\beta}$ results of volatility factor

IFEmoving_window

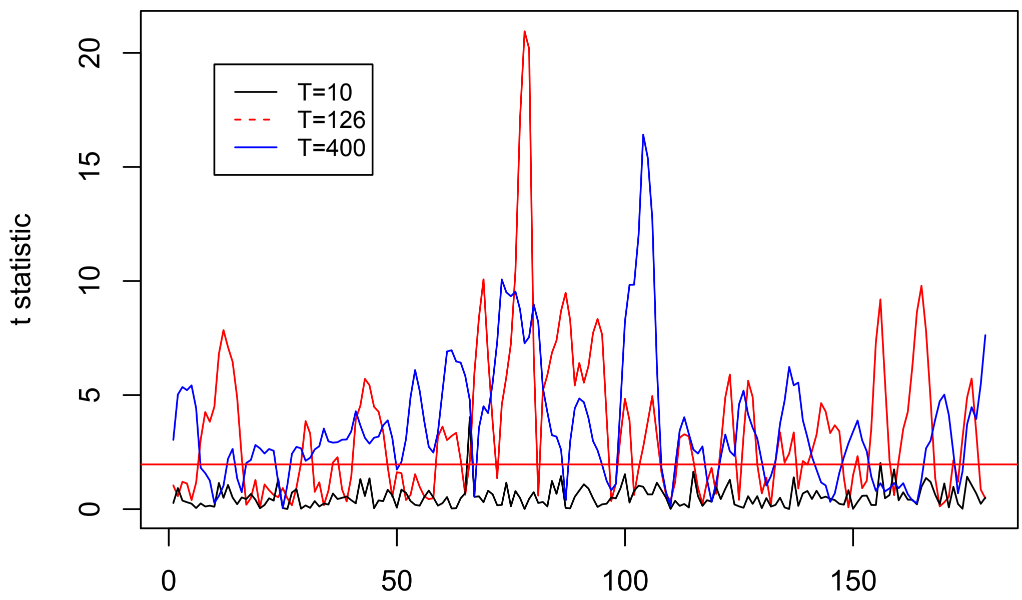

Figure 10: t-statistic of size factor

IFEmoving_window

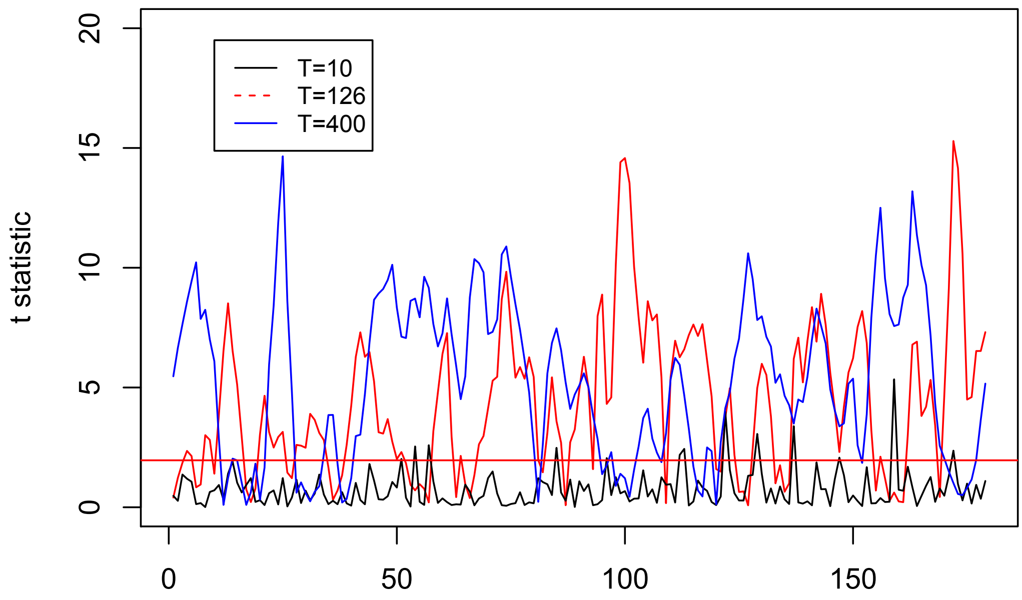

Figure 11: t-statistic of value factor

IFEmoving_window

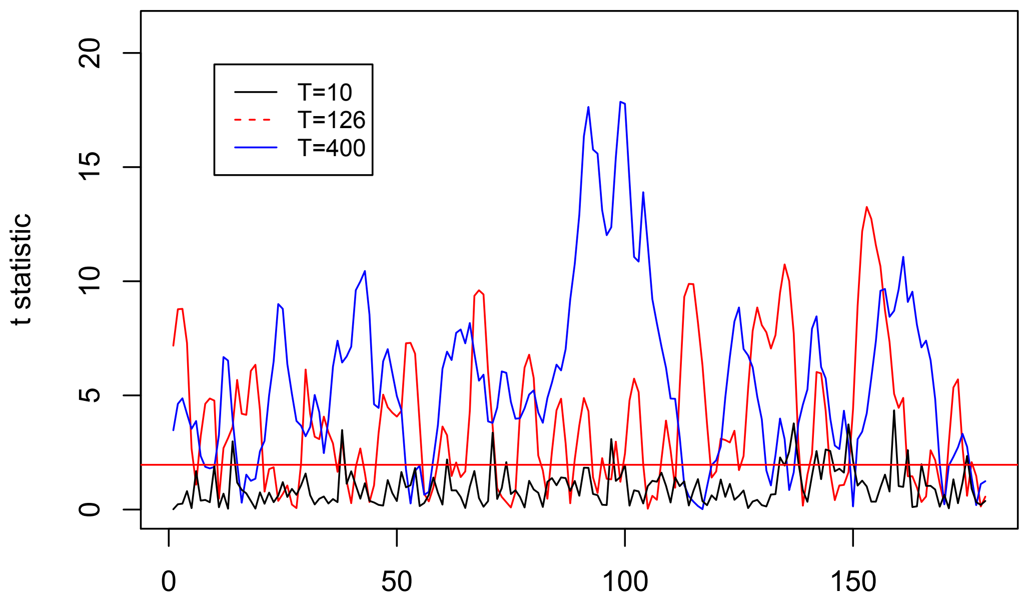

Figure 12: t-statistic of profitability factor

IFEmoving_window

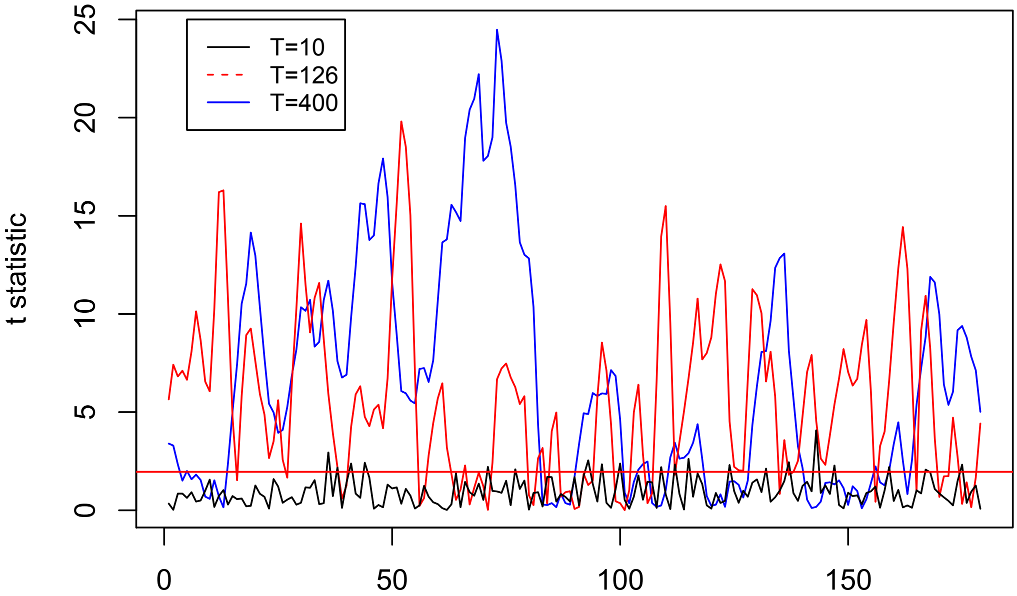

Figure 13: t-statistic of momentum factor

IFEmoving_window

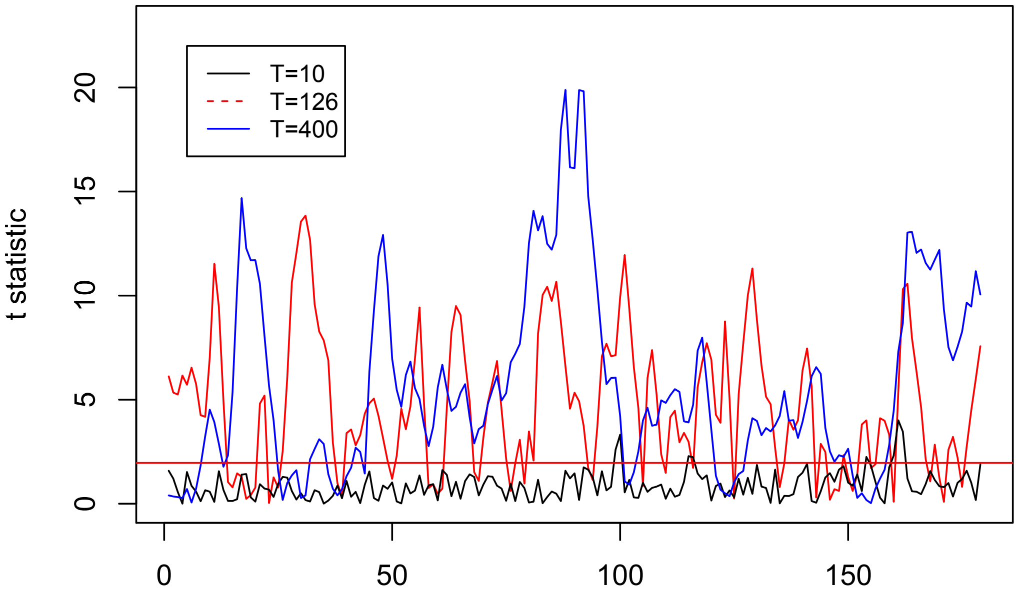

Figure 14: t-statistic of volatility factor

IFEmoving_window

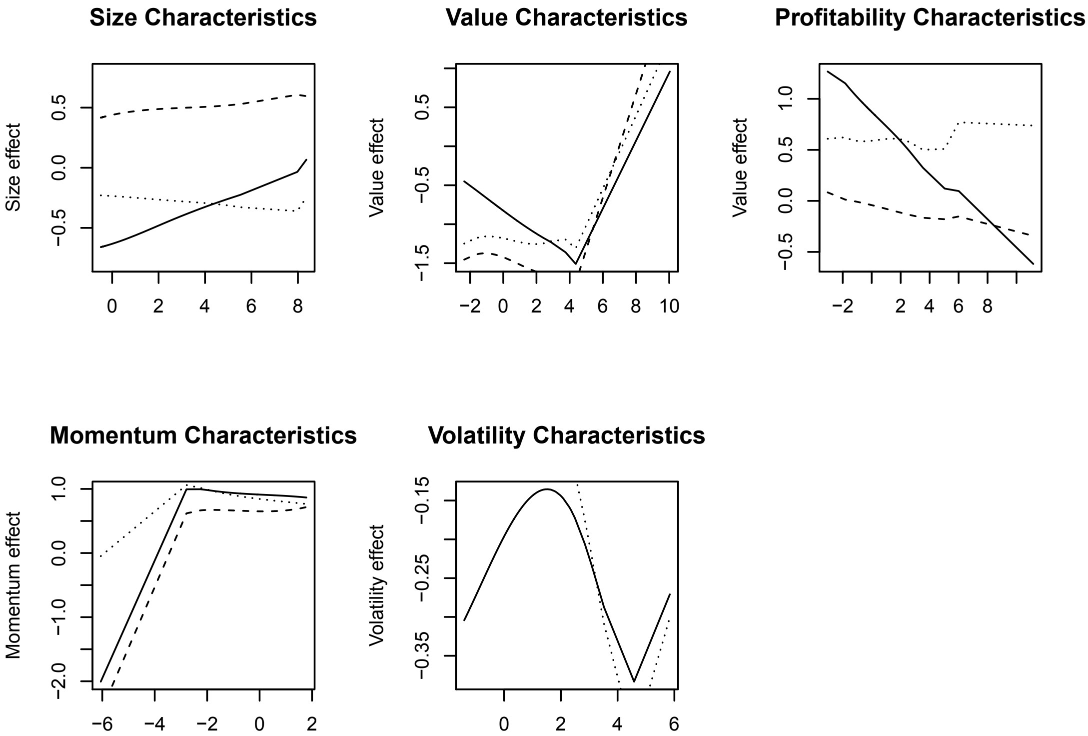

Figure 15: Estimated results of additive functions g_{kl} of the first 126 trading days, $l = 1, \dots, 11$. The solid, dashed and dotted lines are related to the first, second and third factors

QIFEestimation

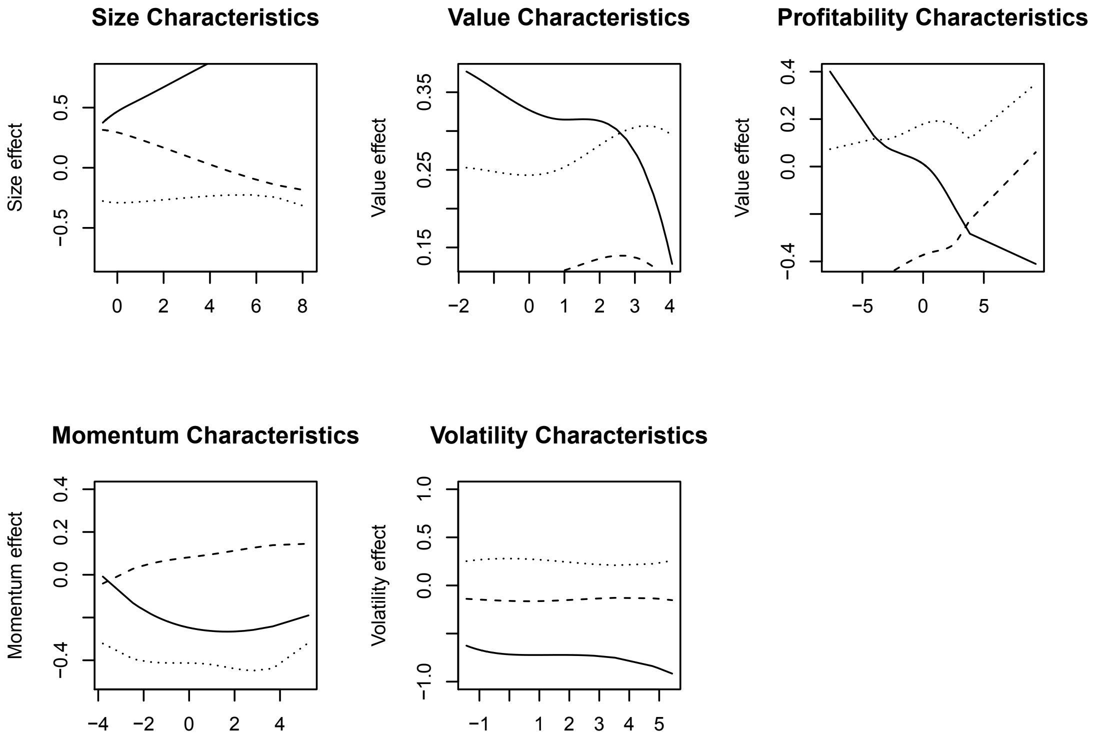

Figure 16: Estimated results of additive functions g_{kl} of the last 126 trading days, $l = 1, \ldots, 11$. The solid, dashed and dotted lines are related to the first, second and third factors.

IFEestimation

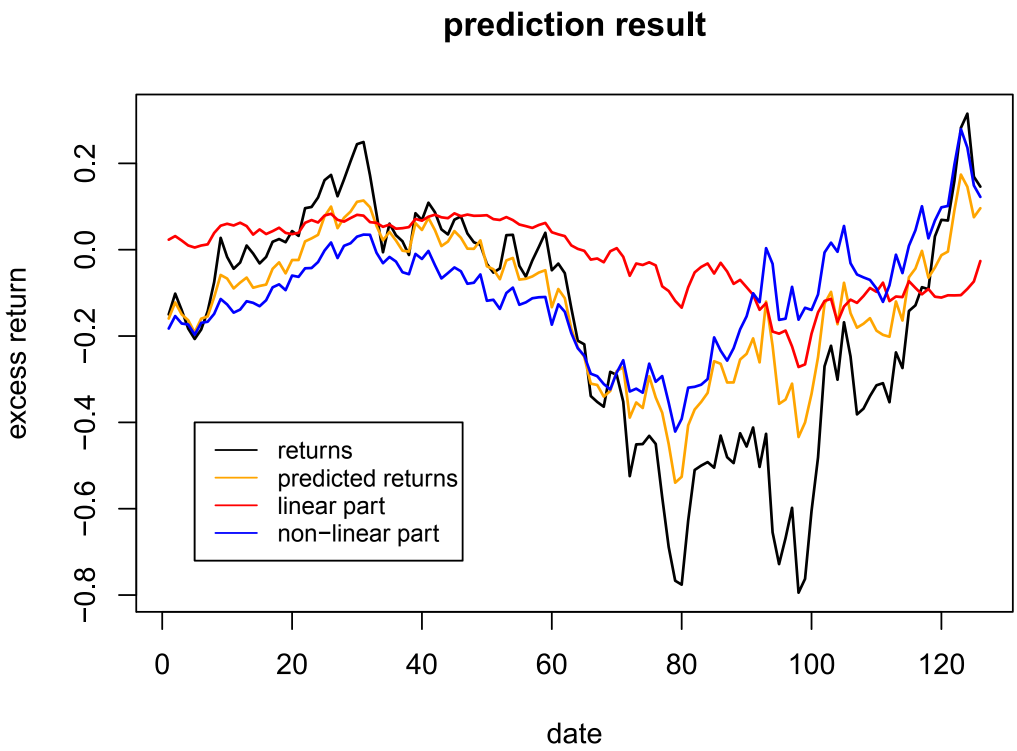

Figure 17: prediction results of one sample of the first window of 126 trading days

QIFEestimation

5 Conclusions

The results in the previous section shows the firm-specific characteristics has not only linear but also nonlinear significant impact on the excess returns. The portfolio grouping method suggested in Fama and French (2016) is considered the equivalent to applying nonlinear transformation on the covariates, thus a significant linear relationship can be obtained. Our proposed framework allows for both estimated linear coefficients which can be easy to explain and additive nonlinear part with unknown factors entering the underlying model governing the relationship between returns and characteristics. This latter feature of our proposed framework bridges the gap between portfolio sorts and cross-sectional regressions and will allow investors to select a portfolio intuitively based on firm-specific characteristics.

But in order to apply this model in a dynamic environment, we need to further develop a dynamic panel data model with dynamic factors, which can be challenging in estimation in the future.

Another challenge is to discuss the application of the methods in different assumption of the prior distributions of the data with their initial values.

The application of this model can be not only on stock market but also on prediction of the bond returns, mutual fund returns or even crypto-curencies.

I hope this method could help investors in forming smart-β investing strategy and in portfolio selection.

References

Bai, J. (2009): "Panel data models with interactive fixed effects," *Econometrica*, 77, 1229–1279.

Banz, R. W. (1981): "The relationship between return and market value of common stocks," *Journal of financial economics*, 9, 3–18.

Basu, S. (1983): "The relationship between earnings' yield, market value and return for NYSE common stocks: Further evidence," *Journal of financial economics*, 12, 129–156.

Bollerslev, T., R. F. Engle, and J. M. Wooldridge (1988): "A capital asset pricing model with time-varying covariances," *Journal of political Economy*, 96, 116–131.

Carhart, M. M. (1997): "On persistence in mutual fund performance," *The Journal of finance*, 52, 57–82.

Connor, G., M. Hagmann, and O. Linton (2012): "Efficient semiparametric estimation of the Fama–French model and extensions," *Econometrica*, 80, 713–754.

Connor, G. and O. Linton (2007): "Semiparametric estimation of a characteristic-based factor model of common stock returns," *Journal of Empirical Finance*, 14, 694–717.

Fama, E. F. and K. R. French (1993): "Common risk factors in the returns on stocks and bonds," *Journal of financial economics*, 33, 3–56.

——— (2016): "Dissecting anomalies with a five-factor model," *The Review of Financial Studies*, 29, 69–103.

Fama, E. F. and J. D. MacBeth (1973): "Risk, return, and equilibrium: Empirical tests," *Journal of political economy*, 81, 607–636.

Fan, J., Y. Liao, and W. Wang (2016): "Projected principal component analysis in factor models," *Annals of statistics*, 44, 219.

LETTAU, M. AND S. LUDVIGSON (2001): "Resurrecting the (C) CAPM: A cross-sectional test when risk premia are time-varying," *Journal of Political Economy*, 109, 1238–1287.

LINTNER, J. (1965): "Security prices, risk, and maximal gains from diversification," *The journal of finance*, 20, 587–615.

ROLL, R. AND S. A. ROSS (1980): "An empirical investigation of the arbitrage pricing theory," *The Journal of Finance*, 35, 1073–1103.

ROSENBERG, S. A., M. T. LOTZE, L. M. MUUL, S. LEITMAN, A. E. CHANG, S. E. ETTINGHAUSEN, Y. L. MATORY, J. M. SKIBBER, E. SHILONI, J. T. VETTO, ET AL. (1985): "Observations on the systemic administration of autologous lymphokine-activated killer cells and recombinant interleukin-2 to patients with metastatic cancer," *New England journal of medicine*, 313, 1485–1492.

SHARPE, W. F. (1964): "Capital asset prices: A theory of market equilibrium under conditions of risk," *The journal of finance*, 19, 425–442.

Declaration of Authorship

I hereby confirm that I have authored this Master's thesis independently and without use of others than the indicated sources. All passages which are literally or in general matter taken out of publications or other sources are marked as such.

Berlin, February 08, 2019